LIML and k-class

liwzhang

January 2023

1 Endogenous Regressors

X is exogenous for β if $\mathbb{E}[Xe]=0$. Partition $X=(X_1,X_2)$ with dimensions (k_1,k_2) so that X_1 contains the exogenous regressors and X_2 contains the endogenous regressors. The structural equation is

$$Y = X_1'\beta_1 + X_2'\beta_2 + e. \tag{1.1}$$

The assumptions regarding the regressors and regression error are

$$\begin{aligned}\mathbb{E}\left[X_1 e\right] &= 0\\ \mathbb{E}\left[X_2 e\right] &\neq 0.\end{aligned}$$

2 Instruments

The $\ell \times 1$ random vector Z is an instrumental variable for (1.1) if

$$\mathbb{E}[Ze] = 0 \tag{2.1}$$

$$\mathbb{E}\left[ZZ'\right] > 0 \tag{2.2}$$

$$\operatorname{rank}\left(\mathbb{E}\left[ZX'\right]\right) = k \tag{2.3}$$

Regressors X_1 satisfy condition (2.1) and thus should be included as instrumental variables. They are therefore a subset of the variables Z. Notationally we make the partition

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ Z_2 \end{pmatrix} \begin{matrix} k_1 \\ \ell_2 \end{matrix} \tag{2.4}$$

Here, $X_1 = Z_1$ are the included exogenous variables and Z_2 are the excluded exogenous variables. That is, Z_2 are variables which could be included in the equation for Y (in the sense that they are uncorrelated with e) yet can be excluded as they have true zero coefficients in the equation. We say that the model is just-identified if $\ell = k$ and over-identified if $\ell > k$.

What variables can be used as instrumental variables? From the definition $\mathbb{E}[Ze]=0$ the instrument must be uncorrelated with the equation error, meaning that it is excluded from the structural equation as mentioned above. From the rank condition (2.3) it is also important that the instrumental variables be correlated with the endogenous variables X_2 after controlling for the other exogenous variables $Z_1 = X_1$. These two requirements are typically interpreted as requiring that the instruments be determined outside the system for (Y, X_2), causally determine X_2, but do not causally determine Y except through X_2.

3 Reduced Form

The reduced form is the relationship between the endogenous regressors X_2 and the instruments Z. A linear reduced form model for X_2 is

$$X_2 = \Gamma' Z + u_2 = \Gamma'_{12} Z_1 + \Gamma'_{22} Z_2 + u_2 \tag{3.1}$$

The $\ell \times k_2$ coefficient matrix Γ is defined by linear projection:

$$\Gamma = \mathbb{E}\left[ZZ'\right]^{-1} \mathbb{E}\left[ZX_2'\right] \tag{3.2}$$

This implies $\mathbb{E}\left[Zu_2'\right] = 0$. The projection coefficient (3.2) is well defined and unique under (2.2). We also construct the reduced form for Y. Substitute (3.1) and (2.4) into (1.1) to obtain

$$\begin{aligned} Y &= Z_1'\beta_1 + \left(\Gamma'_{12} Z_1 + \Gamma'_{22} Z_2 + u_2\right)' \beta_2 + e & (3.3)\\ &= Z_1'\lambda_1 + Z_2'\lambda_2 + u_1 & (3.4)\\ &= Z'\lambda + u_1 & (3.5) \end{aligned}$$

where

$$\begin{aligned} \lambda_1 &= \beta_1 + \Gamma_{12}\beta_2 & (3.6)\\ \lambda_2 &= \Gamma_{22}\beta_2 & (3.7)\\ u_1 &= u_2'\beta_2 + e. & (3.8) \end{aligned}$$

We can also write

$$\lambda = \bar{\Gamma}\beta \tag{3.9}$$

where

$$\bar{\Gamma} = \begin{bmatrix} I_{k_1} & \Gamma_{12} \\ 0 & \Gamma_{22} \end{bmatrix}.$$

Together, the reduced form equations for the system are

$$\begin{aligned} Y &= \lambda' Z + u_1 \\ X_2 &= \Gamma' Z + u_2. \end{aligned}$$

or

$$\begin{bmatrix} Y \\ X_2 \end{bmatrix} = \begin{bmatrix} \lambda_1' & \lambda_2' \\ \Gamma'_{12} & \Gamma'_{22} \end{bmatrix} Z + u \tag{3.10}$$

where $u = (u_1, u_2)$. The least squares estimators of (3.2) and (3.5) are

$$\widehat{\Gamma} = \left(\sum_{i=1}^{n} Z_i Z_i'\right)^{-1} \left(\sum_{i=1}^{n} Z_i X_{2i}'\right)$$
$$\widehat{\lambda} = \left(\sum_{i=1}^{n} Z_i Z_i'\right)^{-1} \left(\sum_{i=1}^{n} Z_i Y_i\right)$$

4 Limited Information Maximum Likelihood

Anderson and Rubin (1949) derived the maximum likelihood estimator for the joint distribution of $\vec{Y} = (Y, X_2)$ known as LIML.

This estimator is called "limited information" because it is based on the structural equation for Y combined with the reduced form equation for X_2. If maximum likelihood is derived based on a structural equation for X_2 as well, then this leads to what is known as full information maximum likelihood (FIML). The advantage of the LIML approach relative to FIML is that the former does not require a structural model for X_2, and thus allows the researcher to focus on the structural equation of interest - that for Y. We do not describe the FIML estimator as it is not commonly used in applied econometrics. While the LIML estimator is less widely used among economists than 2SLS it has received a resurgence of attention from econometric theorists. To derive the LIML estimator recall the definition $\vec{Y} = (Y_1, X_2)$ and the reduced form (3.10)

$$\begin{aligned}\vec{Y} &= \begin{bmatrix} \lambda_1' & \lambda_2' \\ \Gamma_{12}' & \Gamma_{22}' \end{bmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} + u \\ &= \Pi_1' Z_1 + \Pi_2' Z_2 + u \end{aligned} \tag{4.1}$$

where $\Pi_1 = \begin{bmatrix} \lambda_1 & \Gamma_{12} \end{bmatrix}$ and $\Pi_2 = \begin{bmatrix} \lambda_2 & \Gamma_{22} \end{bmatrix}$. The LIML estimator is derived under the assumption that u is multivariate normal. Define $\gamma' = \begin{bmatrix} 1 & -\beta_2' \end{bmatrix}$. From (3.7) we find

$$\Pi_2 \gamma = \lambda_2 - \Gamma_{22} \beta_2 = 0.$$

Thus the $\ell_2 \times (k_2 + 1)$ coefficient matrix Π_2 in (4.1) has deficient rank. Indeed, its rank must be k_2 since Γ_{22} has full rank.

This means that the model (4.1) is precisely the reduced rank regression model (section 11.11 in Hansen). The MLE for γ is

$$\widehat{\gamma} = \underset{\gamma}{\operatorname{argmin}} \frac{\gamma' \vec{\boldsymbol{Y}}' \boldsymbol{M}_1 \vec{\boldsymbol{Y}} \gamma}{\gamma' \vec{\boldsymbol{Y}}' \boldsymbol{M}_{\boldsymbol{Z}} \vec{\boldsymbol{Y}} \gamma} \tag{4.2}$$

where $\boldsymbol{M}_1 = \boldsymbol{I}_n - \boldsymbol{Z}_1 \left(\boldsymbol{Z}_1' \boldsymbol{Z}_1\right)^{-1} \boldsymbol{Z}_1'$ and $\boldsymbol{M}_{\boldsymbol{Z}} = \boldsymbol{I}_n - \boldsymbol{Z} \left(\boldsymbol{Z}' \boldsymbol{Z}\right)^{-1} \boldsymbol{Z}'$. The minimization (4.2) is sometimes called the "least variance ratio" problem.

The minimization problem (4.2) is invariant to the scale of γ (that is, $\hat{\gamma}$ is equivalently the argmin for any c) so normalization is required. A convenient choice is $\gamma' \vec{\boldsymbol{Y}}' \boldsymbol{M}_Z \vec{\boldsymbol{Y}} \gamma = 1$. Using this normalization and the theory of the minimum of quadratic forms (Section A.15 in Hansen) $\hat{\gamma}$ is the generalized eigenvector of $\vec{\boldsymbol{Y}}' \boldsymbol{M}_1 \vec{\boldsymbol{Y}}$ with respect to $\vec{\boldsymbol{Y}}' \boldsymbol{M}_{\boldsymbol{Z}} \vec{\boldsymbol{Y}}$ associated with the smalled generalized eigenvalue. (See Section A.14 for the definition of generalized eigenvalues and eigenvectors.) Computationally this is straightforward. For example, in MATLAB, the generalized eigenvalues and eigenvectors of the matrix $\boldsymbol{A}$ with respect to $\boldsymbol{B}$ is found by the command eig $(\boldsymbol{A}, \boldsymbol{B})$. Once this $\hat{\gamma}$ is found, any other normalization can be obtained by rescaling. For example, to obtain the MLE for β_2 make the partition $\widehat{\gamma}' = \left[\begin{array}{cc} \widehat{\gamma}_1 & \widehat{\gamma}_2' \end{array}\right]$ and set $\widehat{\beta}_2 = -\widehat{\gamma}_2/\widehat{\gamma}_1$.

To obtain the MLE for β_1 recall the structural equation $Y = Z_1'\beta_1 + X_2'\beta_2 + e$. Replacing β_2 with the MLE $\widehat{\beta}_2$ and then apply regression. Thus

$$\widehat{\beta}_1 = \left(\boldsymbol{Z}_1'\boldsymbol{Z}_1\right)^{-1} \boldsymbol{Z}_1' \left(\boldsymbol{Y} - \boldsymbol{X}_2\widehat{\beta}_2\right) \tag{4.3}$$

These solutions are the MLE for the structural parameters β_1 and β_2.

There is an alternative (and traditional) expression for the LIML estimator. Define the minimum obtained in (4.2)

$$\widehat{\boldsymbol{\kappa}} = \min_{\gamma} \frac{\gamma'\overrightarrow{\boldsymbol{Y}}'\boldsymbol{M}_1\overrightarrow{\boldsymbol{Y}}\gamma}{\gamma'\overrightarrow{\boldsymbol{Y}}'\boldsymbol{M}_Z\overrightarrow{\boldsymbol{Y}}\gamma} \tag{4.4}$$

which is the smallest generalized eigenvalue of $\overrightarrow{\boldsymbol{Y}}'\boldsymbol{M}_1\overrightarrow{\boldsymbol{Y}}$ with respect to $\overrightarrow{\boldsymbol{Y}}'\boldsymbol{M}_Z\overrightarrow{\boldsymbol{Y}}$. The LIML estimator then can be written as

$$\widehat{\beta}_{\text{liml}} = \left(\boldsymbol{X}'\left(\boldsymbol{I}_n - \widehat{\kappa}\boldsymbol{M}_{\boldsymbol{Z}}\right)\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\left(\boldsymbol{I}_n - \widehat{\boldsymbol{\kappa}}\boldsymbol{M}_{\boldsymbol{Z}}\right)\boldsymbol{Y}\right). \tag{4.5}$$

Expression (4.5) does not simplify computation (since $\widehat{\kappa}$ requires solving the same eigenvector problem that yields $\widehat{\beta}_2$). However (4.5) is important for the distribution theory. It also helps reveal the algebraic connection between LIML, least squares, and 2SLS.

The estimator (4.5) with arbitrary κ is known as a k-class estimator of β. While the LIML estimator obtains by setting $\kappa = \widehat{\kappa}$, the least squares estimator is obtained by setting $\kappa = 0$ and 2SLS is obtained by setting $\kappa = 1$. It is worth observing that the LIML solution satisfies $\widehat{\kappa} \geq 1$.

When the model is just-identified the LIML estimator is identical to the IV and 2SLS estimators. They are only different in the over-identified setting. (One corollary is that under just-identification and normal errors the IV estimator is MLE.) For inference it is useful to observe that (4.5) shows that $\widehat{\beta}_{\text{liml}}$ can be written as an IV estimator

$$\widehat{\beta}_{\text{liml}} = \left(\widetilde{\boldsymbol{X}}'\boldsymbol{X}\right)^{-1}\left(\widetilde{\boldsymbol{X}}'\boldsymbol{Y}\right)$$

using the instrument

$$\widetilde{\boldsymbol{X}} = \left(\boldsymbol{I}_n - \widehat{\boldsymbol{\kappa}}\boldsymbol{M}_Z\right)\boldsymbol{X} = \left(\begin{array}{c} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 - \widehat{\boldsymbol{\kappa}}\widehat{\boldsymbol{U}}_2 \end{array}\right)$$

where $\widehat{\boldsymbol{U}}_2 = \boldsymbol{M}_Z\boldsymbol{X}_2$ are the (reduced-form) residuals from the multivariate regression of the endogenous regressors X_2 on the instruments Z. Expressing LIML using this IV formula is useful for variance estimation.

The LIML estimator has the same asymptotic distribution as 2SLS. However, they have quite different behaviors in finite samples. There is considerable evidence that the LIML estimator has reduced finite sample bias relative to 2SLS when there are many instruments or the reduced form is weak. However, on the other hand LIML has wider finite sample dispersion.

We now derive the expression (4.5). Use the normalization $\gamma' = \left[\begin{array}{cc} 1 & -\beta_2' \end{array}\right]$ to write (4.2) as

$$\widehat{\beta}_2 = \underset{\beta_2}{\operatorname{argmin}} \frac{\left(\boldsymbol{Y} - \boldsymbol{X}_2\beta_2\right)'\boldsymbol{M}_1\left(\boldsymbol{Y} - \boldsymbol{X}_2\beta_2\right)}{\left(\boldsymbol{Y} - \boldsymbol{X}_2\beta_2\right)'\boldsymbol{M}_{\boldsymbol{Z}}\left(\boldsymbol{Y} - \boldsymbol{X}_2\beta_2\right)}.$$

The first-order-condition for minimization is $2/\left(\boldsymbol{Y}-\boldsymbol{X}_2\widehat{\beta}_2\right)'\boldsymbol{M_Z}\left(\boldsymbol{Y}-\boldsymbol{X}_2\widehat{\beta}_2\right)$ times

$$
\begin{aligned}
0 &= X_2'M_1\left(Y-X_2\widehat{\beta}_2\right)-\frac{\left(Y-X_2\widehat{\beta}_2\right)'M_1\left(Y-X_2\widehat{\beta}_2\right)}{\left(Y-X_2\widehat{\beta}_2\right)'M_Z\left(Y-X_2\widehat{\beta}_2\right)}X_2'M_Z\left(Y-X_2\widehat{\beta}_2\right) \\
&= X_2'M_1\left(Y-X_2\widehat{\beta}_2\right)-\widehat{\kappa}X_2'M_Z\left(Y-X_2\widehat{\beta}_2\right)
\end{aligned}
$$

using definition (4.4). Rewriting,

$$
\boldsymbol{X}_2'\left(\boldsymbol{M}_1-\widehat{\boldsymbol{\kappa}}\boldsymbol{M_Z}\right)\boldsymbol{X}_2\widehat{\beta}_2=\boldsymbol{X}_2'\left(\boldsymbol{M}_1-\widehat{\boldsymbol{\kappa}}\boldsymbol{M}_Z\right)\boldsymbol{Y}. \tag{4.6}
$$

Equation (4.5) is the same as the two equation system

$$
\begin{aligned}
\boldsymbol{Z}_1'\boldsymbol{Z}_1\widehat{\beta}_1+\boldsymbol{Z}_1'\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 &= \boldsymbol{Z}_1'\boldsymbol{Y} \\
\boldsymbol{X}_2'\boldsymbol{Z}_1\widehat{\beta}_1+\left(\boldsymbol{X}_2'\left(\boldsymbol{I}_n-\widehat{\boldsymbol{\kappa}}\boldsymbol{M_Z}\right)\boldsymbol{X}_2\right)\widehat{\beta}_2 &= \boldsymbol{X}_2'\left(\boldsymbol{I}_n-\widehat{\boldsymbol{\kappa}}\boldsymbol{M_Z}\right)\boldsymbol{Y}.
\end{aligned}
$$

The first equation is (4.3). Using (4.3), the second is

$$
\boldsymbol{X}_2'\boldsymbol{Z}_1\left(\boldsymbol{Z}_1'\boldsymbol{Z}_1\right)^{-1}\boldsymbol{Z}_1'\left(\boldsymbol{Y}-\boldsymbol{X}_2\widehat{\beta}_2\right)+\left(\boldsymbol{X}_2'\left(\boldsymbol{I}_n-\widehat{\boldsymbol{\kappa}}\boldsymbol{M_Z}\right)\boldsymbol{X}_2\right)\widehat{\beta}_2=\boldsymbol{X}_2'\left(\boldsymbol{I}_n-\widehat{\boldsymbol{\kappa}}\boldsymbol{M_Z}\right)\boldsymbol{Y}
$$

which is (4.6) when rearranged. We have thus shown that (4.5) is equivalent to (4.3) and (4.6) and is thus a valid expression for the LIML estimator.